ITEA Journal 2010; 31: 324–327

VV&A of M&S in the DoD: Moving from Art to Science

John Diem
Associate Director,
Modeling and Simulation Coordination Office, Alexandria, Virginia
Amy Henninger, Ph.D.
Institute for Defense Analyses, Alexandria, Virginia
Simone Youngblood
Johns Hopkins University, Laurel, Maryland

Many a statistician or serious student of statistics will recall his or her introduction to the great George Box quote: "*All models are wrong; some are useful*." Some of us first saw it in the response surfaces text written by Box and Draper (1987), but the quote is actually attributed to Box in one of his much earlier texts (Box 1979; Box and Draper 1987).

We know from Dr. Box's good teachings that to improve the credibility of using a regression model the modeler should prove a number of assumptions (e.g., independence in observations, distribution of residuals is normal around 0) and should test that the model does not significantly overfit or underfit the data (i.e., lack-of-fit tests). While the exact tests used for these measures do not precisely port to the validation of modeling and simulation (M&S), the goal of reporting the observation is a healthy one: Treat the validation of Department of Defense (DoD) M&S more as a science and less as an art.

This article reports on a number of efforts intended to provide the policy infrastructure and some initial tools to help move the DoD in that direction. Specifically, we report on recent changes to verification, validation, and accreditation (VV&A) policy (DODI 5000.61); recent and ongoing activities to advance VV&A (e.g., reporting standards, methodologies in light of risk, VV&A planning and reporting tools); and a series of practical application efforts being conducted with the test and evaluation (T&E) community to fully assess the viability of these evolutions in policy, methodology, standards, and tools. This "VV&A Campaign Plan" is being managed by the Office of Secretary of Defense's Modeling & Simulation Coordination Office (M&S CO) and its technical lead, the Applied Physics Lab at the Johns Hopkins University; however, at conclusion, it will have involved managers and practitioners from across the acquisition and T&E communities.

Department of Defense Instruction 5000.61 (DoDI 5000.61)

In 2007, the DoD M&S Steering Committee embarked on a rewrite of the policy guiding VV&A in the Department, DoDI 5000.61, entitled "DoD Modeling and Simulation Verification, Validation, and Accreditation (VV&A)" (DoD 2009). That revision was completed and vetted in late 2009. The policy overall has been in existence for over 13 years. This document defines policy, assigns responsibilities, prescribes procedures, and establishes common terminology relative to VV&A across the Department. Initially issued in April 1996, the instruction was reissued in May 2003 and, most recently, in December 2009. The most significant changes to the current version of the instruction were:

- modifying the document format to align with updated format requirements for DoD issuances;
- streamlining the document by:
  - pulling the responsibilities section to an enclosure,
  - focusing the procedure section on documentation requirements only,
  - synthesizing documentation procedures by eliminating duplication of information requirements,
  - retaining only essential definitions;
- modifying the document to reference MIL-STD-3022 (DoD 2008), the DoD Standard Practice for the Documentation of VV&A for Models and Simulation.

DoDI 5000.61 provides high-level or "umbrella" policy to which component-based (e.g., the Services; the Director, Operational Test and Evaluation) and organizational (e.g., the Navy's Commander Operational Test and Evaluation Force) policies align. As defined in DoDI 5000.61, it is DoD policy that:

- M&S used to support DoD processes, products, and decisions shall undergo V&V throughout their lifecycles and shall be accredited for an intended use;

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2010		**00-00-2010 to 00-00-2010**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
VV&A of M&S in the DoD: Moving from Art to Science	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Modeling and Simulation Coordination Office,1901 N. Beauregard Street, Suite 500,Alexandria,VA,22311-1705**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**4**	

- VV&A results shall be documented and made accessible;
- each DoD component will be the final authority for validating representations of its own forces and capabilities;
- each DoD is authorized to provide VV&A procedures and guidance based on the intended use and risk of use of the M&S.

In addition to the implementation of the policy statements listed above, DoD components have the responsibility to ensure that V&V resources are provided throughout M&S development, modification, or use.

Activities to advance VV&A methodology

The DoD M&S Steering Committee has sponsored several VV&A-focused tasks with the objective of advancing the state of the practice. These tasks include the development of a risk-driven VV&A methodology as well as automated, consistent documentation formats. These tasks are defined below.

Risk-based VV&A (RBA)

The objective of the risk-based VV&A (RBA) methodology task is to optimize VV&A resource use while minimizing the risks of using a model or simulation. Use risk arises from error and uncertainties in the representations of the model or simulation as well as the consequences that arise from a decision predicated on M&S results. While VV&A seeks to provide evidence relative to M&S error and uncertainty, most efforts operate in a resource-constrained environment. As a consequence, VV&A processes must be tailored to optimize resource use within those constraints, to accommodate the alignment of VV&A with the simulation life cycle, and to address the specific priorities of an intended use.

During the initial (planning) phase of the process, the RBA methodology provides a way to tailor while focusing on those aspects of the M&S that carry the greatest risk relative to the intended use. At the back end of the process, the RBA methodology provides a framework in which to articulate risk in terms of the uncertainties and limitations associated with the M&S capabilities.

The RBA task is in the second year of a two-year effort. A report titled "An Approach for Realizing a Risk-Based VV&A (RBA) Methodology" was completed in March 2010.

MIL-STD-3022 and the VV&A Documentation Tool

MIL-STD-3022, "Department of Defense Standard Practice Documentation of Verification, Validation, and Accreditation (VV&A) for Models and Simulations," was developed by the Modeling and Simulation Coordination Office in coordination with the military departments. It establishes templates for the four core products (Accreditation Plan, V&V Plan, V&V Report, and Accreditation Report) of the M&S verification, validation, and accreditation processes. The intent of this standard is to provide consistent documentation that minimizes redundancy and maximizes reuse of information. This promotes a common framework and interfacing capability that can be shared across all M&S programs within the Department of Defense, other government agencies, and allied nations. MIL-STD-3022 was approved as a Military Standard on June 28, 2008.

In addition to the template standard, the U.S. Navy has led the production of a tool (the DoD VV&A Documentation Tool) that automates the MIL-STD-3022 templates. At this time, access to the tool requires a Common Access Card or External Certification Authority.

While these tasks seek to improve the efficiency and effectiveness of VV&A implementation, technical gaps still exist that impact that efficiency and effectiveness. In an attempt to identify and address these gaps, the M&S Steering Committee has implemented a multiprong approach. The first prong addresses known VV&A gaps that impact successful implementation of the RBA methodology such as the development of a V&V techniques catalog and guidance on the development of acceptability criteria. The second prong focuses on the identification of additional gap areas through a Systems Engineering Research Center–led study focused on VV&A practitioners in the field. The third tier focuses on the establishment of an expert subcommittee that can monitor, review, and mature proposed technical advancements.

Gauging the way forward in support of the test and evaluation community

Fiscal year 2010 and 2011 VV&A within M&S CO were originally planned to focus on the evolution of a tiered validation methodology—not dissimilar to the establishment of Technical Readiness Levels for M&S and distributed M&S capabilities. During initial program reviews for this effort and its linkages to the broader set of DoD VV&A activities, it was determined that the current VV&A investments should first be assessed against the needs of the community—before any additional policies or standards or methodologies were crafted. In particular, DoD leadership wanted to assess the viability of VV&A efforts from the perspective of VV&A practitioners—the organizations and individuals who "do" VV&A as opposed to just managing its policy.

The M&S CO and the Johns Hopkins University team guiding these VV&A efforts fully understand that the best policy, tools, methodologies, templates, and standards do not provide a panacea or "magic fix." VV&A activities in support of the T&E community will always, by definition, be challenging because the T&E team is faced with several inherent—and frequently conflicting—facts:

- Test activities involve new or enhanced systems and systems of systems, which typically do not have a rich body of information or data about their performance, design, or employment.
- T&E in the era of "build and field it faster" will put even greater pressure on the team of evaluators, testers, and technologists performing the VV&A activities. The increasing complexity of systems being fielded—and the need to use distributed M&S capabilities fully integrated with the systems/systems of systems under test realistically replicate operational environments or to make up for shortfalls in available personnel and equipment—make failures during VV&A even more visible and expensive.
- The VV&A team for a given T&E activity is typically drawn from several organizations, may be required to use a broad variety of M&S tools—familiar and unfamiliar—to conduct the test events, and is frequently scattered across the country and only converges on the test site during pretest integration and test event.

Over the next 18 months, a series of practical application proofs of principle activities will be conducted with acquisition and testing organizations within all four military departments. The first proof-of-principle event will focus on the application of VV&A early in the systems life cycle—as a key part of the systems engineering process—and will examine verification methodologies and impacts primarily on early developmental and technical testing. The second proof of principle will focus on the application of the VV&A policies, standards, templates, etc. described above to a series of live-virtual-constructive experimentation events led by the Air Force, with the goal of examining the application of VV&A to technically and geographically distributed test environments. The third proof-of-principle activity being sponsored as part of the VV&A Campaign Plan will be the application of the evolved VV&A capabilities to an ongoing series of vehicle survivability tests being conducted primarily by U.S. land forces (Army and Marines).

Use case 1—Systems engineering overview

The first VV&A proof-of-principle activity will involve the Systems Engineering Research Center, a consortium of 18 universities led by the Stevens Institute of Technology. Systems Engineering Research Center researchers have been tasked to look at tools and process that will facilitate the verification of M&S capabilities (stand-alone and federated).

Use case 2—AGILE Fires overview

The second proof-of-principle activity will be led by the Air Forces' Simulation and Analysis Facility and will focus on the development of a use case based on the AGILE Fires project. This use case will specifically address the unique challenges encountered in performing VV&A of a live, virtual, and constructive distributed simulation environment. The use case will emphasize the development of test scenarios and will also examine the integration of a new component into an established distributed environment that has already undergone VV&A. The objective will be to use existing methodologies and strict configuration control to conduct a reverification, revalidation, and reaccreditation of the scenario by assessing only what changed.

The use case would ultimately define a consistent method for performing and documenting the VV&A effort of a system of systems represented in a distributed simulation environment in order to establish confidence in using live, virtual, and constructive representations.

Use case 3—Vehicle survivability testing overview

This effort is very much in its infancy as this article is being written. It is also, by far, the most challenging—in terms of classification, timeliness, and impacts—of the three proofs of principle. Members of the M&S CO VV&A team will gain insights into its evolved VV&A capabilities while also providing direct support to the teams doing survivability analysis.

Closing remarks

We have addressed a number of initiatives intended to assist VV&A practitioners department-wide with a special focus on the T&E community. Clearly, we are on the verge of a paradigm shift as the community moves toward more complex, next-generation testing procedures. And, while the initiatives in this article are designed to help, we understand that they are but a small step forward in a journey of continuous improvements. Future improvements must focus on the use of advanced statistics (experimental design principles) to make the most of every test and machine learning methods to work toward automating some of the VV&A process. Only by moving VV&A from art to science will we be better positioned to serve our warfighters. The T&E community has long been on

the leading edge of tackling this challenge, and we look forward to your continued innovation and commitment in meeting that goal. ❑

JOHN DIEM is the associate director for tools at the Office of the Secretary of Defense (OSD) Modeling and Simulation Coordination Office (M&S CO). He is also the interim chief of Acquisition Modeling and Simulation, supporting the director, Systems Analysis. As the acquisition M&S lead, he currently sponsors a variety of VV&A-related projects that are evolving current VV&A methodologies, tools, and standards to better meet the needs of users and decision makers. Mr. Diem has over 28 years of experience in the simulation, testing, and training community, approximately half of it as a developmental and operational tester employing modeling and simulation capabilities to create realistic test environments for battle command, intelligence, surveillance, reconnaissance, and communications/network systems. Mr. Diem is on detail to OSD from the Army's Operational Test Command, where he was the chief of the Simulation and Integration Division from 2007 to 2009. He has bachelor of art and master of science degrees from Texas A&M University. E-mail: john.diem@osd.mil

DR. AMY HENNINGER recently returned to the Institute for Defense Analyses (IDA) after serving an Intergovernmental Personnel Act detail as associate director (AD) at the Modeling and Simulation Coordination Office (M&S CO). As AD, she managed a portfolio that includes M&S standards; verification, validation, and accreditation (VV&A); and education. She also represented the United States in international panels on VV&A and led international panels on simulation interoperability. At IDA, Dr. Henninger has led studies in distributed computing and shared data services for a variety of offices in OSD. She has written over 45 articles on a variety of science and technology topics (e.g., neural networks, autonomous agents and multiagent systems, artificial intelligence applications, embedded training systems, analytic methods for uncertainty, games for training, performance metrics for intelligent systems) and has been actively engaged in a number of defense community activities. Dr. Henninger holds six university degrees, culminating in a doctor of philosophy degree in computer engineering. E-mail: ahenning@ida.org

SIMONE YOUNGBLOOD served as the Defense Modeling and Simulation Office's VV&A technical director, the focal point for DoD VV&A, from 1997 to 2007. Leveraging an extensive background in simulation development, modification, and application, Ms. Youngblood has been active in the VV&A community for the past 16 years. She contributed to the Navy's VV&A interim policy guidance and was the editor and coordinator of the DoD VV&A Recommended Practices Guide. From 1994 to 1997, she served as chair of the Distributed Interactive Simulation VV&A Subgroup and was co-author and editor of the IEEE Std 1278.4 Recommended Practices Guide for Distributed Interactive Simulation—Verification, Validation, and Accreditation. She has served as chair of the Simulation Interoperability Workshop VV&A Forum from 1997 to the present. During this time she coordinated development of and made major contributions to the IEEE Std 1516.4-2007; IEEE Recommended Practice for Verification, Validation and Accreditation of a Federation; and MIL-STD 3022 Department of Defense Standard Practice on the Documentation of VV&A for M&S. She holds a bachelor of arts degree in mathematics and bachelor of science and master of science degrees in computer science. E-mail: simone.youngblood@jhuapl.edu

References

Box, G. E. P. 1979. Robustness in the strategy of scientific model building. In *Robustness in Statistics*, ed. R. L. Launer and G. N. Wilkinson. New York: Academic Press.

Box, G. E. P., and Norman R. Draper. 1987. *Empirical model-building and response surfaces*. New York: Wiley.

U.S. Department of Defense. 2009. *DoD modeling and simulation verification, validation, and accreditation (VV&A)*. DoDI 5000.61. Washington, DC: Department of Defense.

U.S. Department of Defense. 2008. *Department of Defense standard practice documentation of verification, validation, and accreditation (VV&A) for models and simulations*. MIL-STD-3022, June 28. Washington, DC: Department of Defense.